Estimating the p-values of robust tests for the linear model

Matias Salibian-Barrera
School of Mathematics and Statistics
Carleton University
1125 Colonel By Drive
Room 4302HP
Ottawa, ON – Canada K1S 5B6

Abstract: There are several proposals of robust tests for the linear model in the literature (see, for example, Markatou, Stahel and Ronchetti, 1991). The finite-sample distributions of these test statistics are not known and their asymptotic distributions have been studied under the assumption that the scale of the errors is known, or that it can be estimated without affecting the asymptotic behaviour of the tests. This is in general true when the errors have a symmetric distribution.

Bootstrap methods can, in principle, be used to estimate the distribution of these test statistics under less restrictive assumptions. However, robust tests are typically based on robust regression estimates which are computationally demanding, specially with moderate- to high-dimensional data sets. Another problem when bootstrapping potentially contaminated data is that we cannot control the proportion of outliers that might enter the bootstrap samples. This could seriously affect the bootstrap estimates of the distribution of the test statistics, specially in their tails. Hence, the resulting p-value estimates may be critically affected by a relatively small amount of outliers in the original data.

In this paper we propose an extension of the Robust Bootstrap (Salibian-Barrera and Zamar, 2002) to obtain a fast and robust method to estimate p-values of robust tests for the linear model under less restrictive assumptions.